Intelligent Calibration and Virtual Sensing for Integrated Low-Cost Air Quality Sensors

Zaidan, Martha Arbayani

2020-11-15

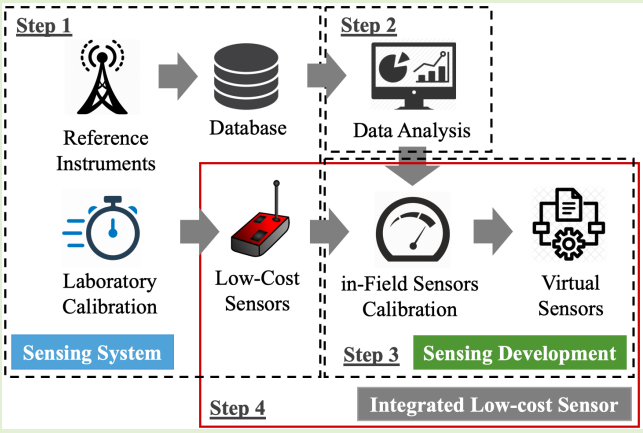
Zaidan , M A , Hossein Motlagh , N , Fung , P L , Lu , D , Timonen , H , Kuula , J , Niemi , J V , Tarkoma , S , Petäjä , T , Kulmala , M & Hussein , T 2020 , ' Intelligent Calibration and Virtual Sensing for Integrated Low-Cost Air Quality Sensors ' , IEEE Sensors Journal , vol. 20 , no. 22 , pp. 13638-13652 . https://doi.org/10.1109/JSEN.2020.3010316

http://hdl.handle.net/10138/321216
https://doi.org/10.1109/JSEN.2020.3010316

Intelligent Calibration and Virtual Sensing for Integrated Low-Cost Air Quality Sensors

Martha Arbayani Zaidan, *Member, IEEE*, Naser Hossein Motlagh, Pak L. Fung, David Lu, Hilkka Timonen, Joel Kuula, Jarkko V. Niemi, Sasu Tarkoma, *Senior Member, IEEE*, Tuukka Petäjä, Markku Kulmala, and Tareq Hussein

***Abstract*—This paper presents the development of air quality low-cost sensors (LCS) with improved accuracy features. The LCS features integrate machine learning based calibration models and virtual sensors. LCS performances are analyzed and some LCS variables with low performance are improved through intelligent field-calibrations. Meteorological variables are calibrated using linear dynamic models. While, due to the non-linear relationship to reference instruments, fine particulate matter ($PM_{2.5}$) are calibrated using non-linear machine learning models. However, due to sensor drifts or faults, carbon dioxide (CO_2) does not present correlation to reference instrument. As a result, the LCS for CO_2 is not feasible to be calibrated. Hence, to estimate the CO_2 concentration, mathematical models are developed to be integrated in the calibrated LCS, known as a virtual sensor. In addition, another virtual sensor is developed to demonstrate the capability of estimating air pollutant concentrations, e.g. black carbon, when the physical sensor devices are not available. In our paper, calibration models and virtual sensors are established using corresponding reference instruments that are installed on two reference stations. This strategy generalizes the models of calibration and virtual sensing which then allows LCS to be deployed in field independently with a high accuracy. Our proposed methodology enables scaling-up accurate air pollution mapping appropriate for smart cities.**



***Index Terms*—Air quality, low-cost sensors, calibration, virtual sensors, machine learning.**

I. Introduction

AIR pollution is a worldwide problem having impacts on both local and global scales. According to the World Health Organization (WHO), air pollution causes 7 million deaths every year, with 4.2 million attributed to exposure to outdoor air pollution [1], [2]. In urban areas, air pollutant is a mixture between local emissions (such as anthropogenic emissions) and those being transported over a regional scale (such as mineral dust). Typically, air pollution measurements are performed by monitoring stations network, which consists of accurate air quality monitoring setup. Such monitoring station network is expensive that make it difficult to densely deploy them within urban areas [3]. Therefore, new methods are needed to provide high spatial resolution of air quality monitoring without losing the accuracy. Recently, low-cost sensors (LCS) have evolved as a promising solution with affordable price [4]–[6].

However, data quality of the LCS remains a major issue that hinders the widespread of LCS implementation. Therefore, many research on LCS have dedicated their studies on developing in-field sensor calibration [7]. For example, LCS calibrations were carried out for improving the sensing variables of Carbon monoxide (CO) [8], Carbon dioxide (CO_2) [9], Nitrogen dioxide (NO_2) [8], Sulfur dioxide (SO_2) [8], Ozone (O_3) [8], [9], particulate matter smaller than 10 μm (PM_{10}) and particulate matter smaller than 2.5 μm ($PM_{2.5}$) [10], [11].

Manuscript received June 18, 2020; accepted July 12, 2020. Date of publication July 20, 2020; date of current version October 16, 2020. This work was supported in part by the MegaSense Program at the University of Helsinki, in part by the City of Helsinki Innovation Fund (UrbanSense), in part by the European Union through the Urban Innovative Action Healthy Outdoor Premises for Everyone under Project UIA03-240, in part by the Helsinki Center for Data Science (HiDATA) Program within the Helsinki Institute for Information Technology (HIIT), in part by the Academy of Finland Centre of Excellence in Atmospheric Science under Grant 307331, in part by the NanoBiomass under Grant 307537, in part by the ACTRIS Finland under Grant 328616, in part by ACTRIS CF under Grant 329274, in part by the Regional Innovations and Experimentations Funds AIKO Governed by the Helsinki Regional Council through the Project HAQT under Grant AIKO014, in part by the Megasense Program, in part by the European Commission through ACTRIS2 under Grant 654109, in part by ACTRIS-IMP under Grant 871115, in part by the SMart URBan Solutions for Air Quality, Disasters, and City Growth under Grant 689443, in part by the ERA-NET-Cofund, University of Helsinki (ACTRIS-HY), in part by the European Research Council via Advanced Grant ATM-GTP under Grant 742206, and in part by the Academy Professor Project of Markku Kulmala. The associate editor coordinating the review of this article and approving it for publication was Prof. Okyay Kaynak. *(Corresponding author: Martha Arbayani Zaidan.)*

Please see the Acknowledgment section of this article for the author affiliations.

Digital Object Identifier 10.1109/JSEN.2020.3010316

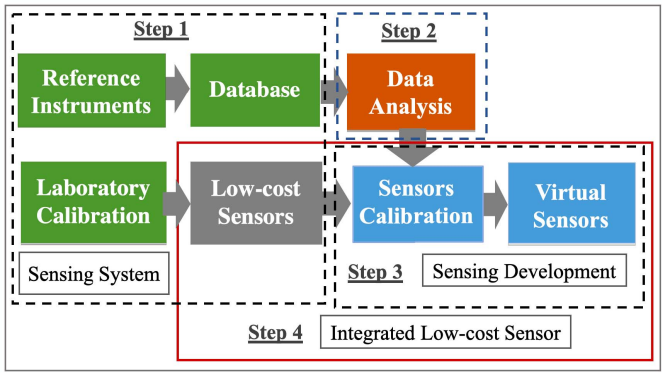


Fig. 1. The methodology of integrating sensors calibration and virtual sensing.

In addition to developing LCS calibration methods, proxies have gained research momentum in air pollution research. A proxy can be defined as a mathematical model that estimates an air pollutant using other available measured variables [12]. The deployment of air pollutant proxy is beneficial to forecast air pollution level [13], to fill missing data from air quality database, and to substitute instruments that are typically expensive and complex in operations [14]. Proxies have been developed to estimate different air pollutant variables such as PM_{10} [15], [16], $PM_{2.5}$ [15]–[17], CO [16], [18], NO_2 [16], [19], SO_2 [16], [20] and O_3 [14], [16], [21]. With regard to CO_2 and BC, several proxies were also developed. For example, Deleawe [22] predicted the level of CO_2 using proxies. The study in [23] aims to demonstrate the use of machine learning methods in predicting ambient CO_2. Other research also attempt to develop proxies for estimating BC indoors and outdoors using regression analysis [24], [25] and machine learning methods [12], [26], [27].

Nevertheless, air pollutant proxies can also be integrated into LCS. Indeed, LCS sometimes cannot be calibrated, because the sensors' drifts and errors lead to missing data or poor measurements performance. In this situation, calibration methods may not be the best solution for improving LCS accuracy. An alternative solution is to develop virtual sensors. *Virtual sensors* can be defined as air pollutant proxies which can be integrated into LCS by using inputs from LCS measurements after calibration. Indeed, virtual sensors can be developed alongside the deployment of LCS.

The integration of LCS calibration and virtual sensors can be defined as *integrated LCS*. Figure 1 depicts the process of developing integrated LCS which proposes a promising solution for two key research questions: (1) how to calibrate LCS in the field and (2) how to extend the operation of LCS to monitor additional air quality indicators that are not measured directly by these LCS. To answer these two questions, in this paper we suggest a novel method of *integrated LCS* embedded with the features of intelligent calibration and virtual sensors. We demonstrate the success of this method by utilizing air quality data monitored by two state-of-the-art stations and four LCS.

II. Methodology

Figure 1 illustrates our proposed methodology for integrating LCS calibration and virtual sensors. *Step* 1 presents sensing systems. Here, a database is established using continuous measurements of LCS (which undergo laboratory calibration) and reference instruments installed in reference stations. This step will be described in section III. Data analysis is performed in *Step* 2 to get insights about the environment including air pollutant and meteorological characteristics. LCS data analysis is also carried out to understand the sensors performance in terms of *consistency* between other LCS units, and *accuracy* compared to reference instruments. This step will be explained in Section IV. Based on the data analysis, decisions can be made if LCS require calibrations or virtual sensing. *Step* 3 demonstrates developing calibrations and virtual sensors which will be described in sections V and VI. Finally, *Step* 4 presents the implementation of *integrated LCS*.

We demonstrate the approach of calibration for fine particulate matter $PM_{2.5}$, which is one of the air quality index component [28]. Air quality index refers to pollution degree of air quality in an easy-to-understand way for public. We will demonstrate the approach of virtual sensors for carbon dioxide (CO_2) and black carbon (BC) monitoring. In this case, LCS for CO_2 cannot be calibrated due to sensor drifts or faults whereas LCS for BC is not yet available. Furthermore, these variables are air pollutants that are harmful to human as well as environment [29], [30] and they are also known to have key impacts on the climate change [31].

In a nutshell, contributions of our methodology, presented in this paper include installing four identical LCS on two highly accurate reference stations over 15 months for data collection; capturing seasonal variations; allowing LCS to be validated between different units (*cross-units validation*) and between different sites (*cross-sites validation*); enabling analysis of LCS *consistency* and *accuracy*, leading to decision making in establishing calibrations or virtual sensing. Another contribution pertains to LCS *Sensitivity analysis* that permits investigating a variety of LCS calibrations based on machine learning methods and allowing method selection (by *model generalization*) for sensor calibration and virtual sensors. Finally, we integrate the results of intelligent LCS calibrations and virtual sensors that demonstrate promising performance.

III. Sensing Systems

The sensing systems which are used to perform air quality measurements are shown in Figures 2(a), 2(b), and 2(c). The locations of these sensing systems are shown in Figure 2(d). In the followings, we explain these sensing systems in detail.

A. Reference Monitoring Stations

This work uses two state-of-the-art monitoring stations, located in Helsinki city, as reference monitoring stations. They are called SMEAR and supersite Mäkelänkatu.

1) SMEAR: Station for Measuring Ecosystem-Atmosphere Relations (we denote as $\mathcal{K}$) is a station to measure the relationship between atmosphere and forest in boreal climate zone for research and scientific exploration [33]. In our study, we use that data at SMEAR III, which is an accurate reference urban air quality monitoring station. This station is located at Kumpula campus of the University of Helsinki in the front

(a) Mäkelänkatu reference station [32]. (b) SMEAR III reference station in Kumpula. (c) Low-cost sensors installed in SMEAR III. (d) Locations of sensing systems.

Fig. 2. Sensing systems used in our experiment. In (d), red triangle, yellow rectangle, and white circles refer to SMEAR III ($\mathcal{K}$), Mäkelänkatu ($\mathcal{M}$), and low-cost sensors, respectively.

open yard and about 4 kilometers north-east from Helsinki center. SMEAR III is also located at 26 meters above sea level on a rocky hill. SMEAR III air quality sensing site is suburban and distinct surface covers built, car parking, road, and vegetation areas. The site is situated at about 150 meters from a main street in Kumpula district in Helsinki [34]. The SMEAR III hosts a suite of trace level gas phase pollutant measurements and the eddy covariance systems in the 31 meter mast. The station is accompanied with the gas and aerosol sensors provide data on fluxes of air pollutants.

To conduct our research, we utilize the data obtained from various reference instruments installed in SMEAR III. These reference instruments measure temperature (Platimun resistance thermometer Pt-100), relative humidity (Platimun resistance thermometer and thin film polymer sensor Vaisala DPA500), $PM_{2.5}$ (Thermo TEOM 1405-D), CO_2 (Picarro G1301 Analyzer) and BC (MAAP Thermo Scientific 5012).

2) Mäkelänkatu: Supersite monitoring station (we denote as $\mathcal{M}$), operated by Helsinki Region Environmental Services Authority (HSY), is known as street canyon and located just beside Mäkelänkatu street, which is one of the arterial roads, in Helsinki. The street is lined with apartment buildings. The street consists of six lanes, two rows of trees, two tramlines and two pavements, in a total 42 meters of width. Everyday, different types of vehicles such as cars, busses, truck cross in this street which cause frequent traffic congestion [35]. The volume of traffic causes a high level of pollution including $PM_{2.5}$ and BC. For our study, the reference instruments which we used from this station measure temperature and relative humidity (Vaisala WXT536), $PM_{2.5}$ (Thermo TEOM 1405), CO_2 (LICOR LI-7000) and BC (MAAP Thermo Scientific 5012).

B. Low-Cost Sensors (LCS)

The LCS used in our experiment are developed by Clarity corporation, a company that is based in Berkeley, California, USA. Currently, the company has installed many low-cost sensors in several mega-cities across the world. Figure 2(c) shows Clarity sensors installed in the SMEAR III site, Kumpula district in Helsinki. The weight of the sensor is 450 grams. The input power of the sensor of 5 volts. In our experiment, we used grid electricity for sensor's input power, however, the sensor is also designed to operate by battery. The battery life time of the sensor in case of harvesting solar power take 1-2 years and without harvesting solar power is about 15 days of continuous measurements. The sensors offer sensing meteorological variables including the Temperature (Temp) that uses band-gap technology and Relative Humidity (RH) which uses capacitive technology. The sensors also measure particulate matter (PM) and CO_2 with laser light scattering technology and metal oxide semiconductor technologies, respectively. The sensors underwent laboratory calibration process, by the manufacturer, using Federal Reference Method (FRM) instruments. The sensors are equipped with LTE-4G communication module to transmit the measured data. The transmitted data is also stored in a cloud platform facilitated by Clarity.[1] The platform allows accessing the raw and visualized data. The data can also be downloaded using a user interface accessible by SmartCity WebApp.[2] The measurement frequency of data varies around 16-23 minutes per data point. Two of these LCS were installed on the container at the SMEAR III about 2 meters from the ground level. Other two LCS were installed on Mäkelänkatu air quality sensing station. The LCS located in Mäkelänkatu were installed on the top of container, about 4 meters above the ground level, where sampling inlets of air quality instruments are placed.

C. Data Collection

The data was collected using four LCS and the reference instruments while all LCS located at reference stations from 13 March 2018 to 18 June 2019. We constructed an air quality database (as shown in the first step in Figure 1) by merging data from all of LCS and downloading relevant data from the reference stations [36]. In addition, to establish

[1] smartcity.clarity.io
[2] clarity.io/documents

a synchronized database, due to different time-resolutions between the measurements of LCS and reference instruments, we merged all of the data by averaging them on hourly basis.

There were several short periods of missing data in measurements of LCS as well as reference instruments in the database. The missing data was due to technical problems associated with faults in data acquisition, sensor measurements and communication loss. The longest average of consecutive missing data from all variables in the database is only 0.0022% of the total number of data points. From total data points, the shortest and longest consecutive missing data are 0% and 0.28%, respectively. In such scenario, it is reasonable to interpolate the missing data. In our case, we resort Akima cubic Hermite interpolation method to fill the missing data [37]. This method is effective in preventing excessive local undulations as well as accurately connecting the flat regions between missing data [38].

IV. Data Analysis

This section explains data analysis of the variables of Temp, RH, $PM_{2.5}$, CO_2, and BC. The analysis allows us to develop appropriate calibration models and virtual sensors using these variables. A comprehensive analysis, including histogram, correlation and performance metrics calculations, is carried out on the air quality database to get insights about the environmental conditions in both stations and the characteristics of measured variables as well as LCS.

Figure 3 displays histograms of these five variables obtained from measurements in sites $\mathcal{K}$ and $\mathcal{M}$. Let us recall that the abbreviations $\mathcal{K}$ and $\mathcal{M}$ refer to reference instruments located in SMEAR III (Kumpula) and Mäkelänkatu sites, respectively. The first row of subplot is temperature histogram, where the variable ranges between -20°C and 30°C in the whole year. It can be seen that there is a very small temperature difference between two reference stations, with the median values are 7.49°C and 8.4°C for sites $\mathcal{K}$ and $\mathcal{M}$, respectively. The temperature histograms for two sites are also overlapped between each other, indicating the temperature is almost the same between these two sites. The second row of subplot refers to RH histogram. RH between two sites is slightly different: site $\mathcal{K}$ seems to be more humid than site $\mathcal{M}$.

The last three rows of subplots present histograms for the pollutants. They demonstrate concentrations of $PM_{2.5}$, CO_2 and BC in site $\mathcal{M}$ are frequently higher than site $\mathcal{K}$, especially the median of $PM_{2.5}$ in site $\mathcal{M}$ is almost double than in site $\mathcal{K}$. This takes place because site $\mathcal{M}$ is near pollutant sources as described in the subsection III-A. Nevertheless, from the data analysis and previous studies, Helsinki air quality is relatively good for most of the time compared to many other cities around the world [39]. However, this leads to another challenge in LCS usage. Indeed, LCS is known to be not sensitive to measure very low-concentration of air pollutants [40].

Figure 4 displays a matrix plot generated through the analysis of Pearson correlation coefficient (PCC) in order to understand the relationships between all relevant measured variables (i.e., Temp, RH, $PM_{2.5}$, CO_2 and BC) obtained from

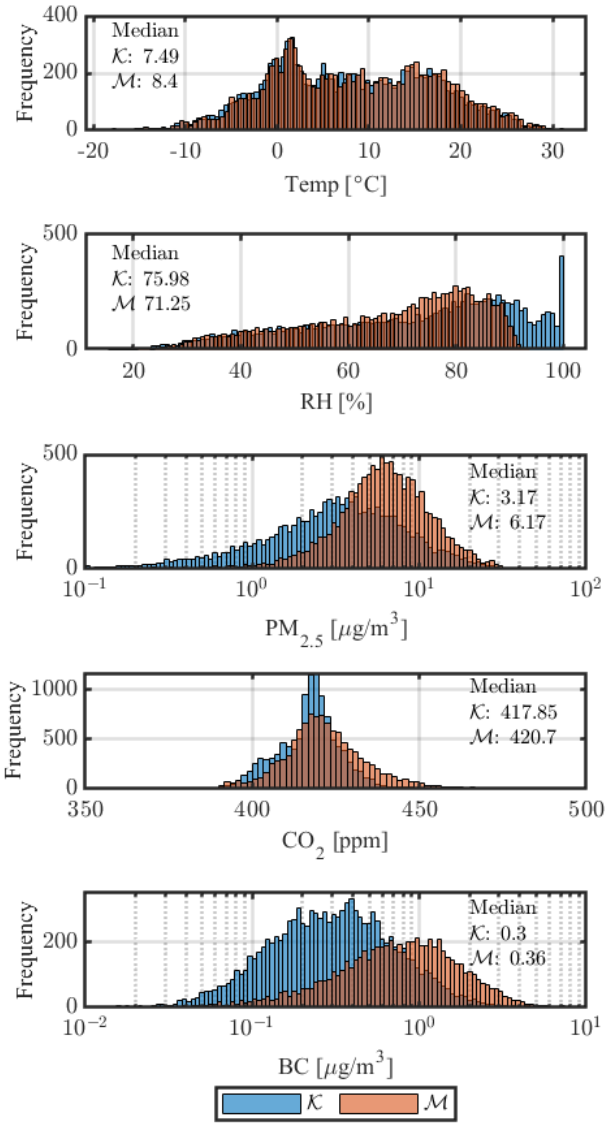


Fig. 3. The histograms of environmental variables used in this study, obtained from reference instruments located in Kumpula $\mathcal{K}$, colored by blue, and Mäkelänkatu $\mathcal{M}$, colored by red.

reference instruments as well as LCS. In Figure 4, the bar plot colour indicates the levels of absolute number of PCC. The color close to dark blue represents very weak correlations whereas the color close to yellow represents strong correlations. The notations Klc and Mlc denote to LCS installed in sites $\mathcal{K}$ and $\mathcal{M}$, respectively. It can also be seen that there are two LCS of each variable installed at each reference station, for example symbolized by Temp ($Klc1$) and Temp ($Klc2$) for units one and two of temperature LCS installed in site $\mathcal{K}$, respectively. Therefore, Figure 4 demonstrates all relationship between the measured variables by LCS (Klc and Mlc) and the reference instruments performed in SMEAR III and Mäkelänkatu sites ($\mathcal{K}$ and $\mathcal{M}$). It can be seen that the correlation levels between the same measured variables of LCS provide almost perfect correlation, except for CO_2. The results demonstrate that all LCS units are consistent if the same units are compared between each other. The matrix plot can also be used to understand the accuracy of LCS by comparing them

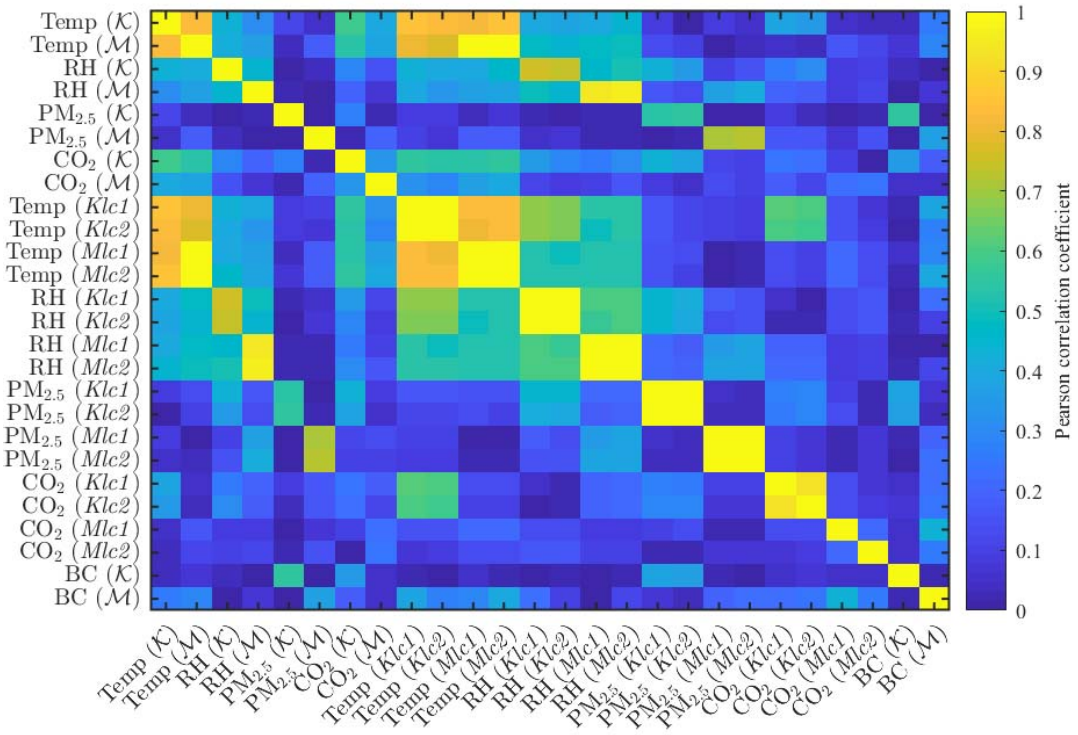


Fig. 4. Correlation analysis between reference instruments and low-cost sensors (LCS) in sites Kumpula $\mathcal{K}$ and Mäkelänkatu $\mathcal{M}$.

TABLE I
THE PERFORMANCE METRICS USED IN THIS STUDY

Performance Metrics	Formulation
Mean Absolute Error	$\text{MAE} = \frac{\sum_{i=1}^{n} \lvert \hat{y}_i - y_i \rvert}{n}$
Root Mean Squared Error	$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}}$
Pearson Correlation Coefficient	$r = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{\hat{y}})\,(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (\hat{y}_i - \bar{\hat{y}})^2}\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}$

TABLE II
THE PERFORMANCE METRICS OF LCS BEFORE FIELD-CALIBRATION AGAINST THE REFERENCE STATIONS

	$\mathcal{K}$			$\mathcal{M}$		
	r	**MAE**	**RMSE**	r	**MAE**	**RMSE**
Temp	0.73	4.024	6.695	0.99	1.113	1.753
RH	0.68	11.102	16.176	0.94	5.556	6.972
$PM_{2.5}$	0.56	8.640	14.74	0.67	7.8	12.54
CO_2	-0.237	14.3	20.135	0.234	18.55	27.712

against the reference instruments. The results will be used and discussed in sections V and VI.

Another data analysis is by calculating several performance metrics. Table I summarizes three performance metrics used for data analysis and performance analysis of calibration as well as virtual sensors. MAE, called the mean absolute error, has a simple interpretation as the average absolute difference between the estimated measurement values and the reference measurement data. RMSE, called root mean squared error, is the standard deviation of the estimated errors. r, also known as PCC, is a measure of the linear correlation between two variables, estimated measurement values and reference measurement data. In the table, the reference measurement data, the mean of the reference measurement data, the estimated data, and the mean of the estimated data are symbolized by y, $\bar{y}$, $\hat{y}$ and $\bar{\hat{y}}$, respectively. The notations of i indicates the index of a data point and n refers to the total number of data points.

Table II presents performance metrics of all variables of LCS in both site, by comparing them with reference instruments. These metrics are used to describe the overall quality of LCS measurements. While the correlation (r) between Temp ($\mathcal{M}$) and Temp (Mlc) are almost perfect ($r = 0.99$), the Table also illustrates imperfect correlations ($r = 0.73$) between Temp ($\mathcal{K}$) and Temp (Klc). Similar results are also obtained for measured RH for both site, i.e. between Mlc and $\mathcal{M}$ ($r = 0.94$) as well as between Klc and $\mathcal{K}$ ($r = 0.68$). The results for Temp and RH in K necessitate the need for calibrations of LCS deployed in site $\mathcal{K}$.

The correlation for $PM_{2.5}$ between K and Klc as well as M and Mlc present lower r values than Temp and RH, that are equal to 0.56 and 0.67, respectively. Since the performance metrics of $PM_{2.5}$ are poorer in general compared to Temp and RH, more complex calibration methods need to be developed.

Finally, the variable CO_2 presents poor correlations between K and Klc as well as M and Mlc with r are equal to -0.237 and 0.234, respectively. This indicates that the measurements of LCS on both sites do not follow the pattern of reference instruments measurements. The reason would be because of sensor drifts or faults. Therefore, the sensor calibrations are not suitable solution for improving the accuracy of CO_2 measurements.

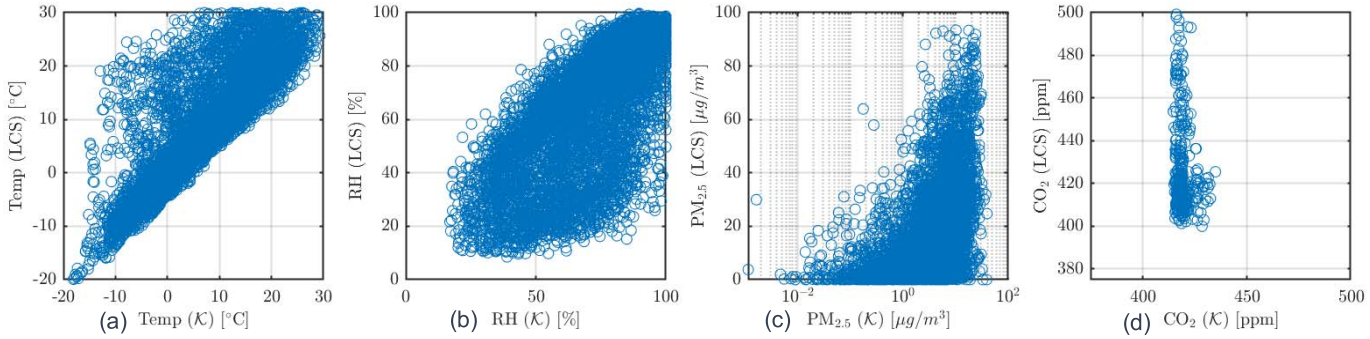


Fig. 5. Scatter plots between low-cost sensors (LCS) and reference instruments in the Kumpula site ($\mathcal{K}$).

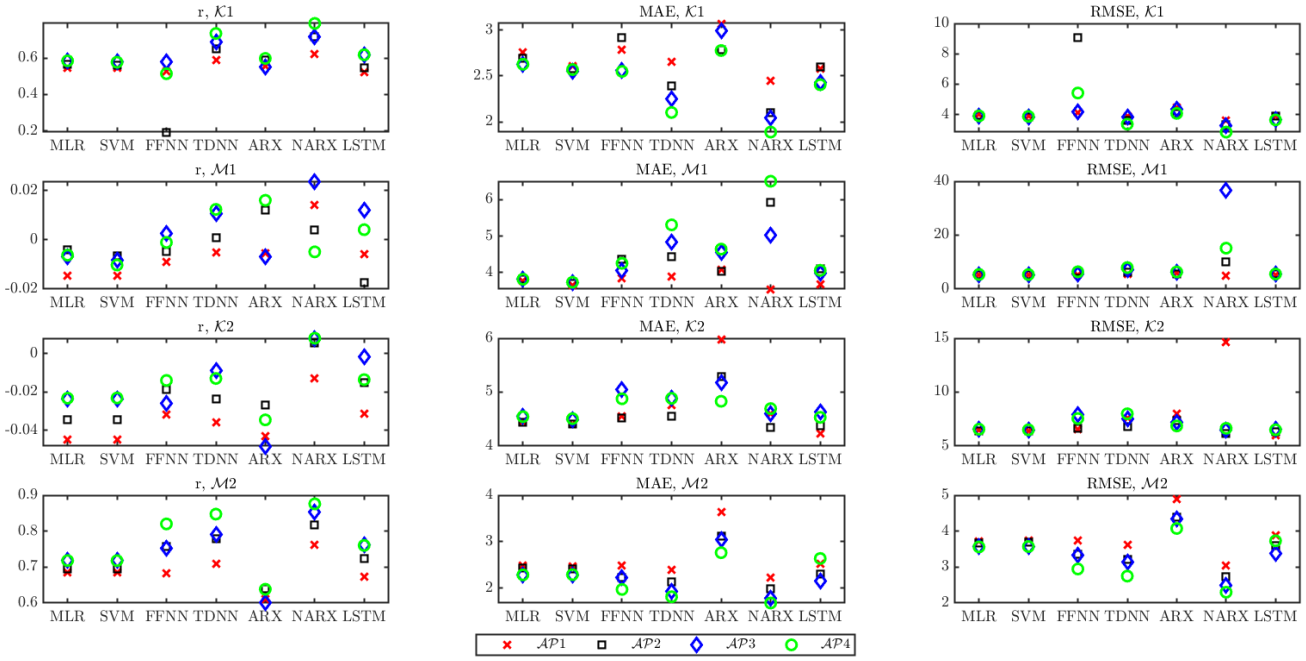


Fig. 6. Sensitivity analysis of different machine learning-based calibration models using four $\mathcal{AP}$s for $PM_{2.5}$ concentration for sites $\mathcal{K}$ and $\mathcal{M}$.

V. Sensors Calibration

As shown in Figure 4 and Table II, beside CO_2, all LCS have consistent performance compared with different sensor units at the same measurement site. Nevertheless, there are some LCS measurements which need to be calibrated to meet the data quality at the reference stations. The following subsections describe the development methodology of calibration models for each measured variable.

A. Meteorological Variables

In this subsection, we discuss the calibrations of variables Temp and RH from the LCS measurements. As illustrated in Figure 4 and Table II, LCS measuring these variable which are located in site $\mathcal{M}$, do not require calibration because, according to the performance metrics, the readings of these sensors are reliable. However, the records by LCS for Temp and RH which are deployed in site $\mathcal{K}$ need calibration as they exhibit lower performance compared with the reference instruments.

In Figures 5(a) and 5(b), scatter plots demonstrate that the relationship between the measurements of reference instruments and LCS for Temp and RH in site $\mathcal{K}$ shows linear trends but they do not correlate well. On this basis, we calibrate the variables of these LCS using dynamic linear models. In particular, we resort Autoregressive model with Extra Input (ARX), which has become one of standard tools in linear time-series modelling [41].

Definition 1: ARX is also known as Auto-Regressive models with eXogenous variables. The exogenous variable is referred as the input term. ARX can be defined mathematically as:

$$A(q)y(t) = B(q)u(t - n_k) + e(t) \tag{1}$$

where $A(q)$ *and* $B(q)$ *are given by*:

$$A(q) = 1 + a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a} \tag{2}$$

$$B(q) = b_1 q^{-1-n_k} + \cdots + b_{n_b} q^{-n_b-n_k} \tag{3}$$

where the notation $e(t)$ *is the Gaussian noise. The notations of* a_{n_a} *and* b_{n_b} *are the model parameters. The symbol* n_a *indicates the order of the polynomials of the output* $A(q)$ *whereas the symbols* n_b *and* n_k *represent the order of the polynomials of the input* $B(q)$. q^{-1} *is defined as the shift operator, such as* $u(t-1) = q^{-1}u(t-n_k)$. *The notations of* $u(t)$ *and* $y(t)$ *represent input and output, respectively.*

TABLE III
LCS OF TEMP AND RH AFTER DEPLOYING CALIBRATION METHOD ON SITE KUMPULA ($\mathcal{K}$)

Site:	$\mathcal{K}$	
LCS vars:	Temp	RH
r	0.98	0.90
MAE	1.688	6.430
RMSE	2.285	8.337

In our case, the notation $u(t)$ represents the LCS output of Temp or RH before calibration, whereas the notation $y(t)$ is the outputs of Temp ($\overline{Temp}$) and RH ($\overline{RH}$) after calibration. In order to select the best model configuration, we build a set of models by specifying $n_a = [1, 2, \cdots, 5]$, $nb = [1, 2, \cdots, 10]$ and $nk = [1, 2, \cdots, 5]$. In order to find the best ARX model structure, we compute loss functions for sets of the specified ARX structures using instrumental variable method where the ARX model parameters are estimated using a least-squares method [42]. The ARX calibration models are trained using the data generated from the reference instruments as output, $y(t)$, and LCS ($Klc1$) of Temp/RH as inputs, $u(t)$. The trained ARX is then validated using another LCS unit ($Klc2$). Through this validation, we ensure that the calibration model works on different LCS units. This validation strategy is also known as *cross-units validation* which will also be discussed later.

Table III shows the performance metrics of LCS for both variables after applying the calibration models. The results of r values have improved significantly to 0.98 and 0.90 for Temp and RH, respectively. After calibration, the performance metric of MAE between the reference instruments and LCS, approximately for Temp is 1.688°C and for RH is 6.430%. The performance metric of RMSE are 2.285°C for Temp and 8.337% for RH. These values are much lower than the median values for Temp and RH in site $\mathcal{K}$, which are 8.4°C and 71.25%, respectively (Figure 3). This means that our calibration models have highly improved the measurement accuracy of the Temp and RH for LCS. In addition to having accurate meteorological information, the calibrated Temp and RH of LCS can also be used to further develop calibration models for other variables as well as for developing virtual sensors.

B. $PM_{2.5}$ Concentration

As shown in Figure 4, the measurements of $PM_{2.5}$ for all LCS both sites require calibration. Table II also emphasizes that the r values of $PM_{2.5}$ on both sites are lower than the r values of uncalibrated LCS of Temp and RH (i.e. site $\mathcal{K}$). In addition, Table II explains that MAE with values of 8.640 $\mu g/m^3$ in site $\mathcal{K}$ and 7.8 $\mu g/m^3$ in site $\mathcal{M}$ as well as RMSE with values 8.640 $\mu g/m^3$ in site $\mathcal{K}$ and 12.54 $\mu g/m^3$ in site $\mathcal{M}$ are considered to be unreliable. This is because the values of these performance metrics are much higher than the median values of $PM_{2.5}$ concentrations which are 3.173 $\mu g/m^3$ and 6.168 $\mu g/m^3$, in sites $\mathcal{K}$ and $\mathcal{M}$, respectively. Figure 5(c) shows the scatter plot of $PM_{2.5}$ between the reference instruments and LCS. The $PM_{2.5}$ measurements of LCS do not exhibit linear trends, however, the measurements still demonstrate correlation to the reference instruments.

Based on these reasons, developing calibration models that improve the reading accuracy of $PM_{2.5}$ of LCS is feasible. In practice, recently several calibration based on machine learning models have been developed and tested to improve the accuracy of $PM_{2.5}$ readings. The examples of these calibration models include linear regression (LR) [43], multivariate linear regression (MLR) [9], support vector machine (SVM) [44], feed-forward neural-networks (FFNN) [45], [46], time-delay neural networks (TDNN) and nonlinear autoregressive with exogenous inputs network (NARX) [47].

In our paper, in order to find an optimal calibration model, we implement and test these aforementioned calibration models. We also resort ARX and long-short term memory (LSTM) [48] models. With this investigation, we aim to test, compare and select the best calibration model to improve the accuracy of $PM_{2.5}$ measurements. To establish these calibration models, in training phase, we use the datasets obtained from the reference instruments, while for implementing them we use the output of calibrated LCS. In addition, to estimate $PM_{2.5}$ concentrations with higher accuracy, we perform *sensitivity analysis* to select the best variable combinations. The *sensitivity analysis* aims to establish $PM_{2.5}$ functions ($\mathcal{F}$) consisting of different combination of measured variables by LCS. To this end, we define four approaches as followings:

$$\mathcal{AP}1: \ \overline{PM}_{2.5} = \mathcal{F}(PM_{2.5}) \tag{4}$$
$$\mathcal{AP}2: \ \overline{PM}_{2.5} = \mathcal{F}(\overline{Temp}, \ PM_{2.5}) \tag{5}$$
$$\mathcal{AP}3: \ \overline{PM}_{2.5} = \mathcal{F}(\overline{RH}, \ PM_{2.5}) \tag{6}$$
$$\mathcal{AP}4: \ \overline{PM}_{2.5} = \mathcal{F}(\overline{Temp}, \ \overline{RH}, \ PM_{2.5}) \tag{7}$$

where, $\overline{Temp}$ and $\overline{RH}$ are LCS outputs of Temp and RH after applying calibration models. $PM_{2.5}$ and $\overline{PM}_{2.5}$ are the LCS outputs of $PM_{2.5}$ variable before and after calibrations, respectively.

Figure 6 shows the results of sensitivity analysis of $PM_{2.5}$ calibration using the four approaches ($\mathcal{AP}$s) and different machine learning methods. The symbols $\times$ (red), $\square$ (black), $\diamond$ (blue), and $\circ$ (green) represent approaches 1, 2, 3 and 4, respectively. The results comprise performance metrics of r, MAE and RMSE which are validated on four LCS units. $\mathcal{K}1$ refers when training in site $\mathcal{K}$ and testing in the same site. $\mathcal{K}2$ indicates when training in site $\mathcal{K}$ and testing in site $\mathcal{M}$. $\mathcal{M}1$ describes when training in site $\mathcal{M}$ and testing in site $\mathcal{K}$. $\mathcal{M}2$ refers when training in site $\mathcal{M}$ and testing in site $\mathcal{M}$. Let us also recall that in our experiments, we have used four low-cost sensors, where we called them $Klc1$, $Klc2$, $Mlc1$, and $Mlc2$ as shown in Figure 2(d) and presented in Figure 4. When applying machine-learning methods, we always train using one low-cost sensor and test the model on another low-cost sensors. Table IV summarizes how training and testing are done.

Through the illustration in Table IV, we demonstrate *cross-units validation* and *cross-sites validation*. *Cross-units validation* refers to when a calibration model is trained and tested within the same site between different sensor units.

TABLE IV
TRAINING AND TESTING USING THE DATA FROM FOUR LCS UNITS INSTALLED SITES $\mathcal{K}$ AND $\mathcal{M}$

Reference sensors:	$\mathcal{K}$		$\mathcal{M}$	
Low-cost sensors:	*Klc1*	*Klc2*	*Mlc1*	*Mlc2*
$\mathcal{K}1$	train	test	-	-
$\mathcal{K}2$	train	-	-	test
$\mathcal{M}1$	-	test	train	-
$\mathcal{M}2$	-	-	train	test

For example, for $\mathcal{K}1$ (Table IV), the calibration model is established using $PM_{2.5}$ from unit $Klc1$ and tested on $PM_{2.5}$ from unit $Klc2$ in site $\mathcal{K}$. In addition, *cross-sites validation* refers to when the calibration model is trained in one site and tested in another site. For example, for $\mathcal{K}2$ (Table IV), the calibration model is established using $PM_{2.5}$ from unit $Klc1$ in site $\mathcal{K}$ and tested on $PM_{2.5}$ from unit $Mlc2$ in site $\mathcal{M}$.

Indeed, several conclusions can be drawn from Figure 6. First, *cross-sites validation* does not show promising results. For example, the values of r ($\mathcal{M}1$) for all methods and all approaches are very low, ranging between -0.02 to 0.02. Similarly, the values of MAE ($\mathcal{M}1$) and RMSE ($\mathcal{M}1$) are very large, ranging between 4-6 $\mu g/m^3$, and 0-40 $\mu g/m^3$, respectively. Second, non-linear models such as NARX shows a very good performance for *cross-units validation* but tend to over-fit on *cross-sites validation*. For instance, NARX ($\mathcal{M}2$) depicts very high values for r and very low for MAE and RMSE. However, this method performs very poor on *cross-sites validation*, such as $\mathcal{K}2$ and $\mathcal{M}1$. Other linear methods, such as MLR, seem to perform better than non-linear methods, but in general their performance are also very poor for *cross-sites validation*. These results indicate that all of the presented calibration models are not generalized. They do not function well when they are tested on different sites other than training sites.

In order to cope with the above problem, we need to generalize calibration model. We need to start by selecting one promising result from Figure 6 using metrics from *cross-units validation*. In the figure, it can be seen that non-linear time series models, such as TDNN and NARX, have the best performance for *cross-units validation*. In particular, NARX demonstrates the best performance with the values for r, MAE and RMSE approximately equal to 0.79, 1.890 $\mu g/m^3$ and 2.840 $\mu g/m^3$ for $\mathcal{K}1$ and 0.88, 1.683 $\mu g/m^3$ and 2.301 $\mu g/m^3$ for $\mathcal{M}2$, respectively.

To generalize the model, we train calibration model using the database from two different sites. When dealing with static models and concurrent batches of static data, the data sets can be typically appended together to form one large concurrent batch. However, using the same strategy would cause a discontinuity in the data sequence if time sequences are appended together. Therefore, we create a concurrent set of input and output sequences to be fed into the NARX model.

Definition 2: NARX is known as Non-linear ARX model. This model is a part of recurrent dynamic networks which relates the current value of a time series with the trained and estimated past values of the inputs (exogenous) series. NARX

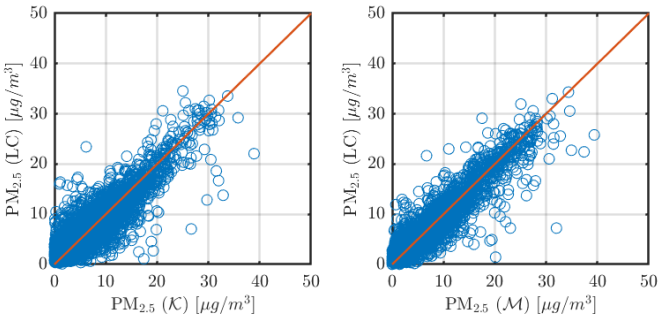


Fig. 7. Scatter plots between reference instruments and calibrated $PM_{2.5}$ LCS for sites $\mathcal{K}$ (left subplot) and $\mathcal{M}$ (right subplot).

can be expressed mathematically as:

$$y(t) = \Psi(y(t-1),\ y(t-2), \cdots,\ y(t-n_y), \\ u(t-1),\ u(t-2), \cdots,\ u(t-n_x)) \quad (8)$$

where $\Psi(.)$ *is a non-linear mapping function of ARX model.* $y(t)$ *is the model output at time* t*. The notations:* $y_{t-1},\ y_{t-2}, \cdots,\ y_{t-n_y}$ *and* $u_{t-1},\ u_{t-2}, \cdots,\ u_{t-n_x}$ *are the past outputs' values and the past inputs' values, respectively. The symbol* n_x *is the number of input delays whereas the symbol* n_y *is the number of output delays. The mapping function* $\Psi(\cdot)$ *is initially unknown and it is approximated during the training process of the prediction.*

In our case, the final implementation uses $\mathcal{AP}4$, so that $y(t)$ is the $PM_{2.5}$ concentration obtained from the reference instrument whereas $u(t)$ is the inputs of LCS of calibrated Temp ($\overline{Temp}$), calibrated RH ($\overline{RH}$) and LCS of $PM_{2.5}$ concentration ($\overline{PM}_{2.5}$). For non-linear function $\Psi(\cdot)$, we use a multilayered neural network. Then, our model is known as NARX neural network, where the internal architecture that performs this approximation is known as Multi-Layer Perceptron (MLP) [41], [49]–[51]. The MLP offers a powerful structure allow to learn any type of continuous nonlinear mapping [52]. In this work, the best NARX architectures are found through grid search, even though most architectures do not lead to significant differences in performance metrics. For all cases, Bayesian regularization backpropagation [53], [54]. is used for NARX parameters' estimation to ensure the model generalization and avoid over-fitting.

Table V shows the results of generalized NARX calibration model trained on both sites $\mathcal{K}$ and $\mathcal{M}$ using the data from the sensor units $Klc1$ and $Mlc2$. The calibration models are tested on the other units, i.e., $Klc2$ and $Mlc1$, located on the different sites. The values of r increase significantly to be 0.90 and 0.94 after calibration. The MAE and RMSE values for both site are about 1 $\mu g/m^3$ and 2 $\mu g/m^3$, respectively. These values are much lower than prior to applying calibrations. The values are also lower than the median values of $PM_{2.5}$ concentration in sites $\mathcal{K}$ and $\mathcal{M}$ (as shown in Figure 3), which are about 3.173 $\mu g/m^3$ and 6.168 $\mu g/m^3$, respectively. The result demonstrates that the generalized calibration model performs well on different units (i.e., *cross-units validation*) and on both sites (i.e., *cross-sites validation*).

Figure 7 displays the scatter plots between the reference instruments and the best calibrated $PM_{2.5}$ LCS in sites $\mathcal{K}$

TABLE V
NARX TRAINED ON BOTH SITES IS THE BEST $PM_{2.5}$ MACHINE LEARNING BASED CALIBRATION MODELS

Train Site:	Sites $\mathcal{K}$ & $\mathcal{M}$	
Test Site:	$\mathcal{K}$	$\mathcal{M}$
r	0.90	0.94
MAE	1.376	1.102
RMSE	2.019	1.607

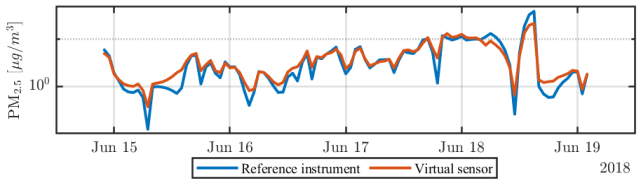


Fig. 8. Time series plot for $PM_{2.5}$ of reference instruments (blue) and calibrated LCS (red) in site $\mathcal{K}$, between 15-19 June 2018.

(left subplot) and $\mathcal{M}$ (right subplot). The outputs of calibrated $PM_{2.5}$ LCS are almost in agreement with the reference red line. Figure 8 shows time-series plot for $PM_{2.5}$ concentration of reference instruments (blue) and calibrated LCS (red) in site $\mathcal{K}$, between 15-19 June 2018. The figure depicts that output of calibrated $PM_{2.5}$ LCS follows the trend of real $PM_{2.5}$ measurements obtained from the reference instruments. In conclusion, the reading of $PM_{2.5}$ after applying the proposed calibrations is considered to be highly accurate and reliable.

C. CO_2 Concentration

As shown in Table II, LCS for CO_2 cannot be calibrated because the performance between the reference instruments and LCS are very poor using all metrics. In addition, in Figure 5(d), the scatter plot between the reference instrument and LCS for variable CO_2 illustrates very poor correlation. This takes place due to the sensors' drifts or faults. Therefore, to improve such poor correlation, in this paper we develop virtual sensors to estimate CO_2 concentration using calibrated LCS outputs. In next section, we discuss our approach for developing virtual sensors.

D. Remarks

NARX has been used extensively in many applications [55]. In air pollution related research, NARX has been used for sensors calibration [47] and air pollutants forecasting [56]. Due to these reasons and our *Sensitivity analysis*, in the rest of our paper, we will use NARX implementation for developing virtual sensors as we used it for calibrating $PM_{2.5}$. We will also establish generalized NARX model which should work across different sites. As the main target of this study is not investigating the algorithms of calibration and virtual sensing, but our main objective is in developing and presenting the methodology of integrated LCS. We state that other machine learning methods may also work better on different cases, datasets and implementation. Nevertheless, our proposed methodology is applicable to any machine learning method as calibration models and virtual sensors.

VI. VIRTUAL SENSORS

The development of virtual sensors are indeed necessary when calibrations are not adequate to provide reliable air quality data or physical sensor devices are not available. Virtual sensors can be developed by applying machine learning methods using reliable and accurate datasets from reference instruments. Once the trained virtual sensors have been established, they can be implemented by using inputs from calibrated LCS outputs. The virtual sensors can be implemented in computing centers known as cloud or even they can be embedded in LCS hardware. In this fashion, virtual sensors can be utilized in fields independently anywhere from reference stations.

Our methodology of integrated LCS calibration and virtual sensors is illustrated in Figure 9. The top sub-figure illustrates the models' development of calibration and virtual sensors. In the development phase, to establish the calibration models (i.e. Temp Cal, RH Cal, $PM_{2.5}$ Cal), we use the measurements data obtained from LCS (before calibration) and the reference instruments. The models of virtual sensors (i.e. CO_2 and BC VS) are established by only using the measurements data obtained from the reference instruments. The established calibration and virtual sensors' models are then deployed in LCS shown in the bottom sub-figure. In this case, LCS measure three variables of Temp, RH and $PM_{2.5}$. In the deployment phase, LCS outputs of Temp and RH are calibrated first using the models of Temp Cal and RH Cal, established previously in the top sub-figure. $PM_{2.5}$ output is then calibrated using the model $PM_{2.5}$ Cal, established in the top sub-figure, by taking additional inputs from calibrated Temp ($\overline{Temp}$), calibrated RH ($\overline{RH}$). Next, all calibrated LCS outputs are fed to the established virtual sensors' models (i.e. CO_2 VS and BC VS) in order to estimate CO_2 and BC concentrations. It is worth noting that the proposed methodology illustrated in Figure 9 is generic, other air pollutants and meteorological variables can also be calibrated, estimated and integrated using this scheme.

In order to demonstrate the usefulness of virtual sensors (as depicted in shown in Figure 9), in this section we present two virtual sensing of variables CO_2 and BC. The first virtual sensing aims to estimate CO_2 concentration when the sensor cannot be calibrated. Second virtual sensing estimates BC concentration when the measurement of variable is not typically available on LCS [12], [57].

In this paper, we use similar NARX implementation to develop these virtual sensors, as described in subsection V-D. To establish virtual sensors, in training phase, we use the datasets from the reference instruments, while in implementation phase we use the output of calibrated LCS. However, virtual sensors can also use outputs of other virtual sensors as their inputs. To estimate CO_2 and BC concentrations using virtual sensors, we define the following functions ($\mathcal{F}$) with different combinations by defining approaches ($\mathcal{AP}$ 5-8):

$$\mathcal{AP}5: \widehat{CO_2} = \mathcal{F}(\overline{Temp}, \overline{RH}, \overline{PM_{2.5}}) \quad (9)$$

$$\mathcal{AP}6: \widehat{BC} = \mathcal{F}(\overline{Temp}, \overline{RH}, \overline{PM_{2.5}}) \quad (10)$$

$$\mathcal{AP}7: \widehat{CO_2} = \mathcal{F}(\overline{Temp}, \overline{RH}, \overline{PM_{2.5}}, \widehat{BC}) \quad (11)$$

$$\mathcal{AP}8: \widehat{BC} = \mathcal{F}(\overline{Temp}, \overline{RH}, \overline{PM_{2.5}}, \widehat{CO_2}) \quad (12)$$

where the notations of $\overline{Temp}$, $\overline{RH}$ and $\overline{PM_{2.5}}$ are the calibrated outputs for Temp, RH and $PM_{2.5}$ from LCS.

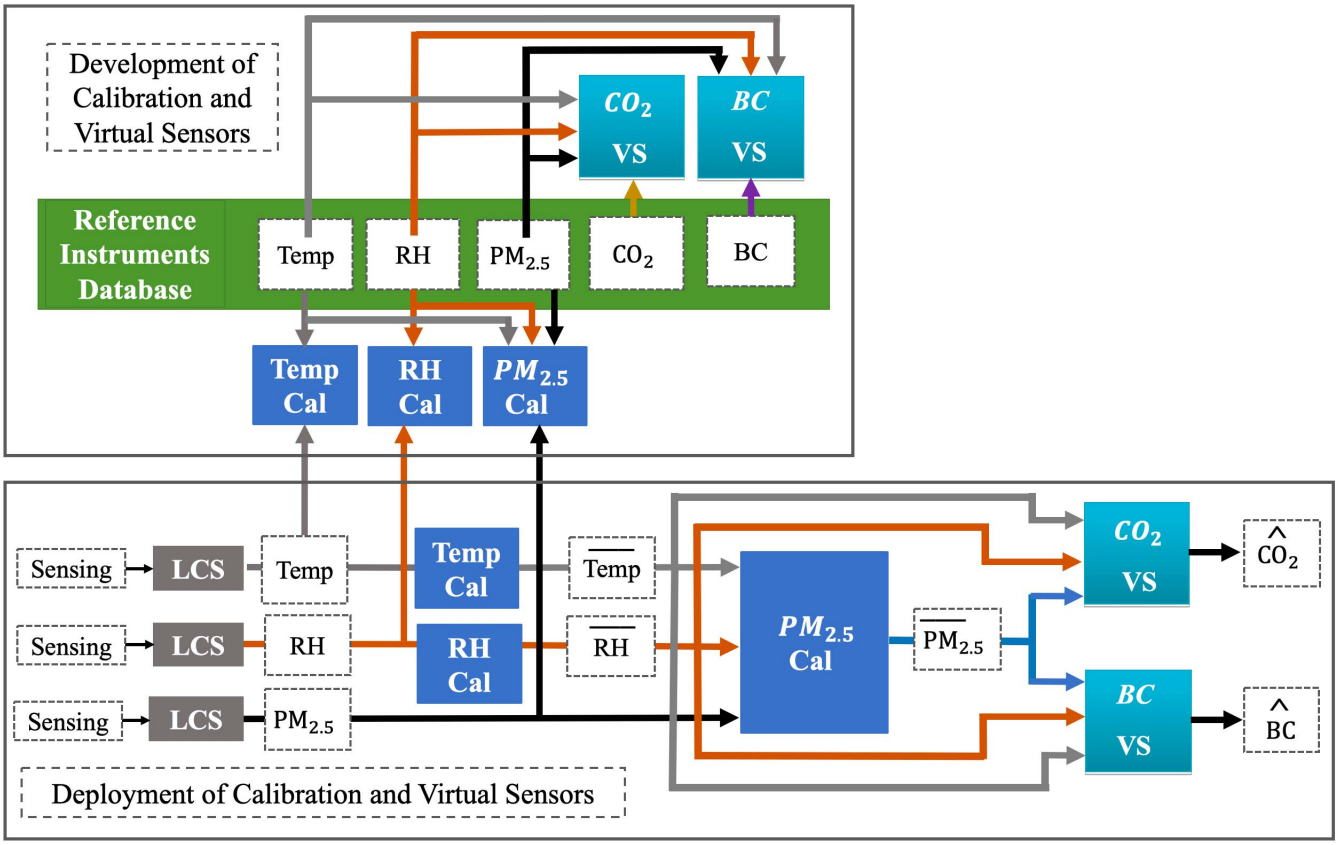


Fig. 9. The development (top sub-figure) and deployment (bottom sub-figure) of integrated low-cost sensors with calibration and virtual sensor features.

TABLE VI
VIRTUAL SENSORS OF CO_2 AND BC USING GENERALIZED NARX MODELS FOR $\mathcal{AP}$S 5-8

	Train Site:	$\mathcal{K}$ - $\mathcal{M}$					
	Test Site:	$\mathcal{K}$			$\mathcal{M}$		
		r	MAE	RMSE	r	MAE	RMSE
CO_2	$\mathcal{AP}5$	0.92	2.935	4.789	0.92	2.498	4.797
	$\mathcal{AP}7$	0.91	3.046	4.933	0.91	2.746	5.115
BC	$\mathcal{AP}6$	0.81	0.219	0.406	0.90	0.141	0.289
	$\mathcal{AP}8$	0.79	0.248	0.427	0.86	0.202	0.342

$\widehat{CO_2}$ and $\widehat{BC}$ are the outputs of virtual sensors of CO_2 and BC, respectively.

As in the case of $PM_{2.5}$ calibration models, we adopt the generalized NARX method by training the model from both sites $\mathcal{K}$ and $\mathcal{M}$. Then, we use the same NARX structure and optimization methods for all approaches. Table VI shows the results for our virtual sensors deployment. All results are generated using the deployment of generalized NARX model.

From the table, we observe that r values of CO_2 are higher than 0.9. The r values for BC virtual sensors are in average of 0.8 and 0.88 at sites $\mathcal{K}$ and $\mathcal{M}$, respectively. The MAE and RMSE values for both virtual sensors are also in agreements with the results of r. The average values of MAE and RMSE of CO_2 virtual sensors for both sites are approximately at 2.81 and 4.91 part per million (ppm). These values are far less than the median values of CO_2 concentrations in sites $\mathcal{K}$ and $\mathcal{M}$ which are equal to 417.88 ppm and 420.70 ppm, respectively. The conclusion is that the values of these metrics can be considered to be excellent.

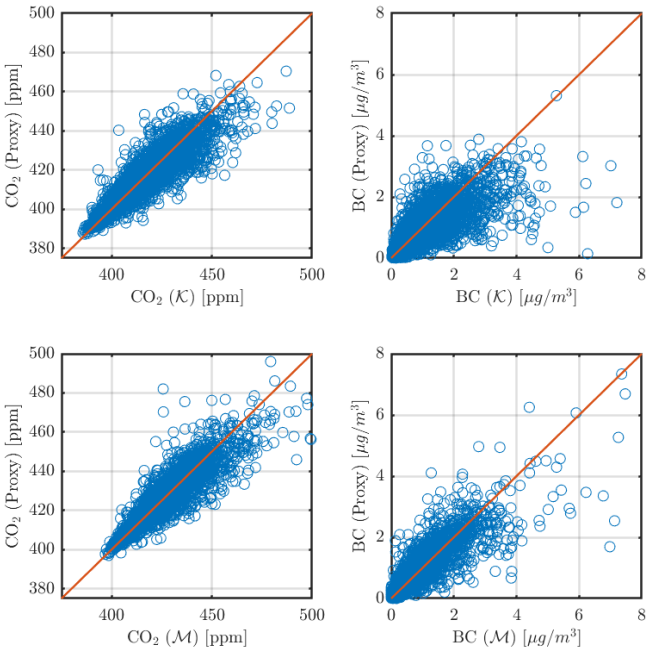


Fig. 10. Scatter plots between reference instruments and CO_2 virtual sensors (left subplots) and BC virtual sensors (right subplots). Top subplots refer to site $\mathcal{K}$ (left subplots) and bottom subplots refer to site $\mathcal{M}$.

Similarly, the mean values of MAE and RMSE of BC virtual sensors for both site are around 0.203 $\mu g/m^3$ and 0.366 $\mu g/m^3$. As the median values of BC concentrations

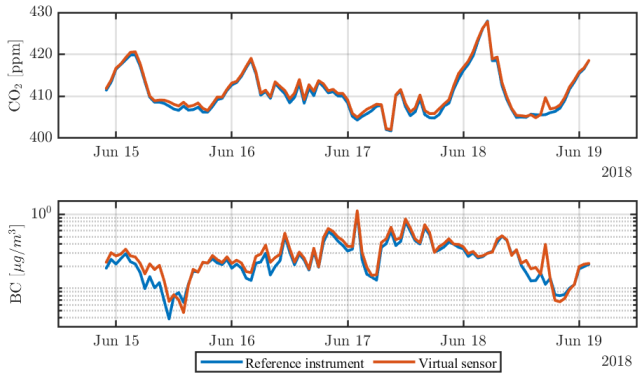


Fig. 11. Time series plot for CO_2 (top subplot) and BC (bottom subplot) of reference instruments (blue) and virtual sensors (red) in site $\mathcal{K}$, between 15-19 June 2018.

in sites $\mathcal{K}$ and $\mathcal{M}$ are 0.3 $\mu g/m^3$ and 0.3550 $\mu g/m^3$, respectively. Therefore, the values of these metrics can be considered to be satisfactory. From the table, almost all performance metrics indicate that $\mathcal{AP}5$ and $\mathcal{AP}6$ are the best virtual sensors for CO_2 and BC, respectively. Whereas, $\mathcal{AP}7$ (for CO_2) and $\mathcal{AP}8$ (for BC) show slightly worse results, because they propagate the modelling errors of LCS calibration and virtual sensors.

Figure 10 shows the scatter plots between the reference instruments and CO_2 virtual sensors (left subplots) and BC virtual sensors (right subplots) in sites $\mathcal{K}$ (top subplots) and $\mathcal{M}$ (bottom subplots). The results for both virtual sensors in both sites are in agreement with the red reference lines. Figure 11 illustrates time-series plots for CO_2 (top subplot) and BC (bottom subplot) of reference instruments (blue) and virtual sensors (red) in site $\mathcal{K}$, between 15-19 June 2018. This figure confirms that both virtual sensors track well the measurement of CO_2 and BC concentrations obtained from reference instruments. These results prove that the virtual sensors implemented on LCS behave similar to the reference instruments. Therefore, the performance of the virtual sensors are satisfactory.

VII. Discussions and Future Works

A. Discussion

The development and applying efficient calibration models and virtual sensors require addressing the deployment, applications and public usage, where we briefly discuss in this section.

1) Deployment: In order to enable the deployment of *integrated LCS* independently from reference instruments, the models of calibrations and virtual sensors can be embedded into LCS hardware or they can also be deployed in any cloud platforms. The former is advantageous in independent deployment without needing Internet and data communication. The latter can be realized through an Internet of Things (IoT) platform by equipping it with a software. This allows model updating mechanism, where the models can be monitored and updated conveniently if necessary. In addition, the LCS drifts and faults can be monitored and maintained. The LCS IoT platform also offers integration between LCS database and local meteorological data centers. As a result, the deployment of LCS can be scaled up realistically in a large quantity to provide high resolution and accurate air quality information.

2) Public Usage: Air quality indexes (AQI) are typically used by government agencies to inform public about the air quality information [58]. AQI gives an overall characterization of actual air quality due to the involvement of several air pollutants in the calculation. In addition, the list of air pollutants is not limited to only $PM_{2.5}$ but include other variables such as SO_2, NO, NO_2, PM_{10} and O_3 concentration [59]. Therefore, to realize the AQI for public usage, our proposed integrated LCS methodology can be implemented for estimating those variables. This can be performed either by adding more LCS sensing (to be calibrated), or by virtual sensors or by integrating both as suggested in this paper. More simplification can also be made by describing the air quality in simple terms (e.g. good, satisfactory, fair, poor, very poor) and an easy-to-understand color scale [60].

3) Applications: In addition to scaling-up air quality map outdoors, reliable LCS is beneficial when they are deployed indoors such as public transports, houses and deep cave mining environments, due to the fact that an individual spends more than 90% of their time indoors in average [61]. Moreover, the proposed integrated LCS is beneficial once upgraded to function as an IoT platform, therefore massive big data can be gathered and analyzed. This helps improving the environment and health related issues [62]–[64], productivity and life-style study, traffic management, and emissions control [65], [66].

B. Future Works

As future works, we plan to embed our proposed methods into the LCS hardware. The field deployment of LCS hardware is challenging due to several factors. First, LCS requires stable and constant electricity for powering the hardware. In order to minimize the energy consumption of LCS, we plan to reduce the physical sensing devices by substituting them with virtual sensors [67]. In addition, we plan to perform experiments to investigate the most efficient data sampling frequency for different measurements. We also aim to investigate energy harvesting methods such as harvesting energy from the surrounding environment for perpetual operation [68].

Second challenge relates to establishing reliable data communication between the LCS and the cloud platforms [69]. In our future studies, we will investigate the network reliability and LCS energy consumption using different communication technologies such as LTE-4G, NB-IoT, and the upcoming 5G networks [70].

In a separate research, we will study the impact of environment on LCS which are deployed in field for long-time. For example, we will investigate the impact of temperature, humidity and the direct sunlight exposed on the LCS to identify the reasons for malfunctioning LCS. The findings can be used to improve functional design and implementation of a complete LCS platform [71].

Calibration and virtual sensors might drift due to environmental changes or sensors degradation. For example, if pollution level deteriorates after the models were developed, these models may not be accurate anymore. Therefore, next we will

develop concept drift detection [72] to monitor if the developed calibration models and virtual sensors are still accurate and reliable. If the models were found to be inaccurate, we will apply online-adaptive models.

Moreover, we plan to integrate the impact of other factors such as wind speed [25], land use, and mobility density to our calibration and virtual sensor models to improve the urban air quality estimation. Last but not least, we will estimate more air pollutants involved in AQI, including SO_2, NO_2, PM_{10}, $PM_{2.5}$, O_3 and CO. This can be realized through LCS hardware extension or incorporating more virtual sensors. Therefore, using massive LCS measurements which will lead to AQI, we aim to provide reliable air quality information with high accuracy and resolution to the public in smart cities.

VIII. Conclusion

In this paper, we present integrated air quality low-cost sensors (LCS) with intelligent calibration and virtual sensing features. This work utilizes more than a year of air quality data sets obtained from reference stations and LCS, installed in those stations. All LCS are consistent, validated between each other in the same measurement site. However, most LCS for measuring meteorological and $PM_{2.5}$ variables have accuracy issues, and thereby they require in-field calibration. Meanwhile, LCS measurements of CO_2 does not present any correlation with the reference instruments due to the sensor drifts or faults. Hence, CO_2 calibration is not feasible. In addition, virtual sensors can also be developed when physical LCS measurements are not available. On these bases, we demonstrate that CO_2 and BC concentrations can be estimated using virtual sensors.

Datasets from the reference stations are used to establish calibration models and virtual sensors. They are then tested on LCS devices. Due to their linear relationship to the reference instruments, Temp and RH are calibrated using dynamic linear models, such as Auto-Regressive models with eXogenous variables (ARX). However, $PM_{2.5}$ needs calibration using more complex models due to non-linear relationship to the reference stations. In this case, through performing sensitivity analysis, calibration models tested in our paper perform well only in the training site, where NARX (Non-linear ARX) is found to be the best calibration model. Therefore, NARX is then generalized by training it with datasets from two reference stations, simultaneously.

The generalization method improves the performance of LCS in measuring $PM_{2.5}$ close to the reference instruments in both sites. Moreover, we continue using generalized NARX to demonstrate development of virtual sensors for variables CO_2 and BC. The results indicate that our virtual sensors estimate CO_2 and BC concentrations with a satisfactory accuracy, without needing real physical sensors for those variables.

In summary, we propose a generic methodology of integrated LCS calibration and virtual sensors, which allows estimating different air pollutants accurately. Any physics-based models or any machine learning methods can be implemented into this methodology. Implementation of the proposed integrated LCS enables independent deployment to produce highly accurate air quality spatio-temporal information.

Acknowledgment

Martha Arbayani Zaidan, Pak L. Fung, Tuukka Petäjä, Markku Kulmala, and Tareq Hussein are with the Institute for Atmospheric and Earth System Research (INAR), University of Helsinki, 00560 Helsinki, Finland (e-mail: martha.zaidan@helsinki.fi; pak.fung@helsinki.fi; tuukka.petaja@helsinki.fi; markku.kulmala@helsinki.fi; tareq.hussein@helsinki.fi).

Naser Hossein Motlagh and Sasu Tarkoma are with the Department of Computer Science, University of Helsinki, 00560 Helsinki, Finland (e-mail: naser.motlagh@helsinki.fi; sasu.tarkoma@helsinki.fi).

David Lu is with Clarity Movements Company, Berkeley, CA 94704 USA (e-mail: david@clarity.io).

Hilkka Timonen and Joel Kuula are with the Finnish Meteorological Institute (FMI), 00560 Helsinki, Finland (e-mail: hilkka.timonen@fmi.fi; joel.kuula@fmi.fi).

Jarkko V. Niemi is with the Helsinki Region Environmental Services Authority (HSY), 00066 Helsinki, Finland (e-mail: jarkko.niemi@hsy.fi).

References

[1] W. H. Organization. (2019). *How Air Pollution is Destroying Our Health*. accessed 2019-12-14. Accessed: Dec. 14, 2019. [Online]. Available: https://www.who.int/airpollution/news-and-events/how-air-pollution-is-d estroying-our-health

[2] P. J. Landrigan *et al.*, "The Lancet Commission on pollution and health," *Lancet*, vol. 391, no. 10119, pp. 462–512, 2018.

[3] E. Lagerspetz *et al.*, "MegaSense: Feasibility of low-cost sensors for pollution hot-spot detection," in *Proc. IEEE 17th Int. Conf. Ind. Informat. (INDIN)*, Helsinki, Finland, Jul. 2019, pp. 23–25.

[4] K. Hu, V. Sivaraman, B. G. Luxan, and A. Rahman, "Design and evaluation of a metropolitan air pollution sensing system," *IEEE Sensors J.*, vol. 16, no. 5, pp. 1448–1459, Mar. 2016.

[5] A. C. Rai *et al.*, "End-user perspective of low-cost sensors for outdoor air pollution monitoring," *Sci. Total Environ.*, vols. 607–608, pp. 691–705, Dec. 2017.

[6] R. Du, P. Santi, M. Xiao, A. V. Vasilakos, and C. Fischione, "The sensable city: A survey on the deployment and management for smart city monitoring," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 2, pp. 1533–1560, 2nd Quart., 2019.

[7] N. H. Motlagh *et al.*, "Toward massive scale air quality monitoring," *IEEE Commun. Mag.*, vol. 58, no. 2, pp. 54–59, Feb. 2020.

[8] C. Borrego *et al.*, "Assessment of air quality microsensors versus reference methods: The EuNetAir joint exercise," *Atmos. Environ.*, vol. 147, pp. 246–263, Dec. 2016.

[9] B. Maag, Z. Zhou, and L. Thiele, "W-air: Enabling personal air pollution monitoring on wearables," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 1, pp. 1–25, 2018.

[10] C.-T. Chiang, "Design of a high-sensitivity ambient particulate matter 2.5 particle detector for personal exposure monitoring devices," *IEEE Sensors J.*, vol. 18, no. 1, pp. 165–169, Jan. 2018.

[11] A. Cavaliere *et al.*, "Development of low-cost air quality stations for next generation monitoring networks: Calibration and validation of $PM_{2.5}$ and PM_{10} sensors," *Sensors*, vol. 18, no. 9, p. 2843, 2018.

[12] M. A. Zaidan, D. Wraith, B. E. Boor, and T. Hussein, "Bayesian proxy modelling for estimating black carbon concentrations using white-box and black-box models," *Appl. Sci.*, vol. 9, no. 22, p. 4976, Nov. 2019.

[13] K. Gu, J. Qiao, and W. Lin, "Recurrent air quality predictor based on meteorology-and pollution-related factors," *IEEE Trans. Ind. Informat.*, vol. 14, no. 9, pp. 3946–3955, Sep. 2018.

[14] M. A. Zaidan *et al.*, "Mutual information input selector and probabilistic machine learning utilisation for air pollution proxies," *Appl. Sci.*, vol. 9, no. 20, p. 4475, Oct. 2019.

[15] H. S. Kim *et al.*, "Development of a daily PM_{10} and $PM_{2.5}$ prediction system using a deep long short-term memory neural network model," *Atmos. Chem. Phys.*, vol. 19, no. 20, pp. 12935–12951, 2019.

[16] H. Maleki, A. Sorooshian, G. Goudarzi, Z. Baboli, Y. T. Birgani, and M. Rahmati, "Air pollution prediction by using an artificial neural network model," *Clean Technol. Environ. Policy*, vol. 21, no. 6, pp. 1341–1352, 2019.

[17] Z. Shang, T. Deng, J. He, and X. Duan, "A novel model for hourly $PM_{2.5}$ concentration prediction based on CART and EELM," *Sci. Total Environ.*, vol. 651, pp. 3043–3052, Feb. 2019.

[18] S. Moazami, R. Noori, B. J. Amiri, B. Yeganeh, S. Partani, and S. Safavi, "Reliable prediction of carbon monoxide using developed support vector machine," *Atmos. Pollut. Res.*, vol. 7, no. 3, pp. 412–418, May 2016.

[19] J. A. Kamińska, "A random forest partition model for predicting NO_2 concentrations from traffic flow and meteorological conditions," *Sci. Total Environ.*, vol. 651, pp. 475–483, Feb. 2019.

[20] A. Masih, "Application of ensemble learning techniques to model the atmospheric concentration of SO_2," *Global J. Environ. Sci. Manage.*, vol. 5, no. 3, pp. 309–318, 2019.

[21] M. Alghamdi *et al.*, "A predictive model for steady state ozone concentration at an urban-coastal site," *Int. J. Environ. Res. Public Health*, vol. 16, no. 2, p. 258, Jan. 2019.

[22] S. Deleawe, J. Kusznir, B. Lamb, and D. J. Cook, "Predicting air quality in smart environments," *J. Ambient Intell. Smart Environ.*, vol. 2, no. 2, pp. 145–154, 2010.

[23] S. Chen, K. Mihara, and J. Wen, "Time series prediction of CO_2, TVOC and HCHO based on machine learning at different sampling points," *Building Environ.*, vol. 146, pp. 238–246, Dec. 2018.

[24] Y. Abu Awad, P. Koutrakis, B. A. Coull, and J. Schwartz, "A spatio-temporal prediction model based on support vector machine regression: Ambient black carbon in three new England States," *Environ. Res.*, vol. 159, pp. 427–434, Nov. 2017.

[25] P. L. Fung *et al.*, "Input-adaptive proxy for black carbon as a virtual sensor," *Sensors*, vol. 20, no. 1, p. 182, Dec. 2019.

[26] C. H. Tang, E. Garshick, S. Grady, B. Coull, J. Schwartz, and P. Koutrakis, "Development of a modeling approach to estimate indoor-to-outdoor sulfur ratios and predict indoor $PM_{2.5}$ and black carbon concentrations for Eastern Massachusetts households," *J. Exposure Sci. Environ. Epidemiol.*, vol. 28, no. 2, pp. 125–130, 2018.

[27] K. Isiugo *et al.*, "Predicting indoor concentrations of black carbon in residential environments," *Atmos. Environ.*, vol. 201, pp. 223–230, Mar. 2019.

[28] Y.-F. Xing, Y.-H. Xu, M.-H. Shi, and Y.-X. Lian, "The impact of $PM_{2.5}$ on the human respiratory system," *J. Thoracic Disease*, vol. 8, no. 1, p. E69, 2016.

[29] N. Caravaggio, S. Caravella, A. Ishizaka, and G. Resce, "Beyond CO_2: A multi-criteria analysis of air pollution in europe," *J. Cleaner Prod.*, vol. 219, pp. 576–586, May 2019.

[30] E. Dons *et al.*, "Transport most likely to cause air pollution peak exposures in everyday life: Evidence from over 2000 days of personal monitoring," *Atmos. Environ.*, vol. 213, pp. 424–432, Sep. 2019.

[31] T. C. Bond *et al.*, "Bounding the role of black carbon in the climate system: A scientific assessment," *J. Geophys. Res., Atmos.*, vol. 118, no. 11, pp. 5380–5552, Jun. 2013.

[32] HSY. (2020). *Helsinki Region Environmental Services*. Accessed: Apr. 4, 2020. [Online]. Available: https://www.hsy.fi/en/

[33] M. Kulmala, "Build a global Earth observatory," *Nature*, vol. 553, no. 7686, pp. 21–23, Jan. 2018.

[34] L. Jarvi *et al.*, "The urban measurement station SMEAR III: Continuous monitoring of air pollution and surface–atmosphere interactions in Helsinki, Finland," *Boreal Environ. Res.*, vol. 14, pp. 86–109, 2009.

[35] R. Hietikko *et al.*, "Diurnal variation of nanocluster aerosol concentrations and emission factors in a street canyon," *Atmos. Environ.*, vol. 189, pp. 98–106, Sep. 2018.

[36] H. Junninen *et al.*, "Smart-SMEAR: On-line data exploration and visualization tool for SMEAR stations," *Boreal Environ. Res*, vol. 14, no. 4, pp. 447–457, 2009.

[37] H. Akima, "A new method of interpolation and smooth curve fitting based on local procedures," *J. ACM*, vol. 17, no. 4, pp. 589–602, Oct. 1970.

[38] H. Akima, "A method of bivariate interpolation and smooth surface fitting based on local procedures," *Commun. ACM*, vol. 17, no. 1, pp. 18–20, Jan. 1974.

[39] K. Teinilä *et al.*, "Concentration variation of gaseous and particulate pollutants in the Helsinki city centre—Observations from a two-year campaign from 2013–2015," *Boreal Environ. Res.*, vol. 24, pp. 115–136, 2019.

[40] N. H. Motlagh *et al.*, "Air quality sensing process using low-cost sensors: Validation by indoor-outdoor measurements," in *Proc. 15th IEEE Conf. Ind. Electron. Appl. (ICIEA)*, Jun. 2020, pp. 1–6.

[41] S. Chen, S. A. Billings, and P. M. Grant, "Non-linear system identification using neural networks," *Int. J. Control*, vol. 51, no. 6, pp. 1191–1214, Jan. 1990.

[42] L. Lennart, *System Identification: Theory for the User*. Upper Saddle River, NJ, USA: Prentice-Hall, 1999, pp. 1–14.

[43] D. Hasenfratz, O. Saukh, S. Sturzenegger, and L. Thiele, "Participatory air pollution monitoring using smartphones," *Mobile Sens.*, vol. 1, pp. 1–5, Apr. 2012.

[44] J. M. Cordero, R. Borge, and A. Narros, "Using statistical methods to carry out in field calibrations of low cost air quality sensors," *Sens. Actuators B, Chem.*, vol. 267, pp. 245–254, Aug. 2018.

[45] Y. Cheng *et al.*, "AirCloud: A cloud-based air-quality monitoring system for everyone," in *Proc. 12th ACM Conf. Embedded Netw. Sensor Syst.*, 2014, pp. 251–265.

[46] Y. Gao *et al.*, "Mosaic: A low-cost mobile sensing system for urban air quality monitoring," in *Proc. 35th Annu. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, Apr. 2016, pp. 1–9.

[47] E. Esposito, S. De Vito, M. Salvato, V. Bright, R. L. Jones, and O. Popoola, "Dynamic neural network architectures for on field stochastic calibration of indicative low cost air quality sensing systems," *Sens. Actuators B, Chem.*, vol. 231, pp. 701–713, Aug. 2016.

[48] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[49] K. S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Trans. Neural Netw.*, vol. 1, no. 1, pp. 4–27, Mar. 1990.

[50] H. T. Siegelmann, B. G. Horne, and C. L. Giles, "Computational capabilities of recurrent NARX neural networks," *IEEE Trans. Syst., Man Cybern. B, Cybern.*, vol. 27, no. 2, pp. 208–215, Apr. 1997.

[51] J. M. P. Menezes and G. A. Barreto, "Long-term time series prediction with the NARX network: An empirical evaluation," *Neurocomputing*, vol. 71, nos. 16–18, pp. 3335–3343, Oct. 2008.

[52] T. Lin, B. G. Horne, P. Tino, and C. L. Giles, "Learning long-term dependencies in NARX recurrent neural networks," *IEEE Trans. Neural Netw.*, vol. 7, no. 6, pp. 1329–1338, 1996.

[53] D. J. C. MacKay, "Bayesian interpolation," *Neural Comput.*, vol. 4, no. 3, pp. 415–447, May 1992.

[54] F. Dan Foresee and M. T. Hagan, "Gauss–Newton approximation to Bayesian learning," in *Proc. Int. Conf. Neural Netw. (ICNN)*, vol. 3, 1997, pp. 1930–1935.

[55] W. R. Jacobs, T. Baldacchino, T. Dodd, and S. R. Anderson, "Sparse Bayesian nonlinear system identification using variational inference," *IEEE Trans. Autom. Control*, vol. 63, no. 12, pp. 4172–4187, Dec. 2018.

[56] E. Pisoni, M. Farina, C. Carnevale, and L. Piroddi, "Forecasting peak air pollution levels using NARX models," *Eng. Appl. Artif. Intell.*, vol. 22, nos. 4–5, pp. 593–602, Jun. 2009.

[57] J. Caubel, T. Cados, and T. Kirchstetter, "A new black carbon sensor for dense air quality monitoring networks," *Sensors*, vol. 18, no. 3, p. 738, Mar. 2018.

[58] A. Rebeiro-Hargrave, N. H. Motlagh, S. Varjonen, E. Lagerspetz, P. Nurmi, and S. Tarkoma, "MegaSense: Cyber-physical system for real-time urban air quality monitoring," in *Proc. 15th IEEE Conf. Ind. Electron. Appl. (ICIEA)*, Jun. 2020, pp. 1–6.

[59] A. Monteiro, M. Vieira, C. Gama, and A. I. Miranda, "Towards an improved air quality index," *Air Qual., Atmos. Health*, vol. 10, no. 4, pp. 447–455, May 2017.

[60] Y. Zheng, F. Liu, and H.-P. Hsieh, "U-Air: When urban air quality inference meets big data," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2013, pp. 1436–1444.

[61] A. Challoner, F. Pilla, and L. Gill, "Prediction of indoor air exposure from outdoor air quality using an artificial neural network model for inner city commercial buildings," *Int. J. Environ. Res. Public Health*, vol. 12, no. 12, pp. 15233–15253, Dec. 2015.

[62] G. Acampora, D. J. Cook, P. Rashidi, and A. V. Vasilakos, "A survey on ambient intelligence in healthcare," *Proc. IEEE*, vol. 101, no. 12, pp. 2470–2494, Dec. 2013.

[63] K. R. Mallires, D. Wang, V. V. Tipparaju, and N. Tao, "Developing a low-cost wearable personal exposure monitor for studying respiratory diseases using metal–oxide sensors," *IEEE Sensors J.*, vol. 19, no. 18, pp. 8252–8261, Sep. 2019.

[64] Y. A. Qadri, A. Nauman, Y. B. Zikria, A. V. Vasilakos, and S. W. Kim, "The future of healthcare Internet of Things: A survey of emerging technologies," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 2, pp. 1121–1167, 2nd Quart., 2020.

[65] M. A. A. Mamun and M. R. Yuce, "Sensors and systems for wearable environmental monitoring toward IoT-enabled applications: A review," *IEEE Sensors J.*, vol. 19, no. 18, pp. 7771–7788, Sep. 2019.

[66] N. H. Motlagh *et al.*, "Indoor air quality monitoring using infrastructure-based motion detectors," in *Proc. IEEE 17th Int. Conf. Ind. Informat. (INDIN)*, Helsinki, Finland, Jul. 2019, pp. 23–25.

[67] C. Sarkar, V. S. Rao, R. Venkatesha Prasad, S. N. Das, S. Misra, and A. Vasilakos, "VSF: An energy-efficient sensing framework using virtual sensors," *IEEE Sensors J.*, vol. 16, no. 12, pp. 5046–5059, Jun. 2016.

[68] F. Touati, A. Galli, D. Crescini, P. Crescini, and A. B. Mnaouer, "Feasibility of air quality monitoring systems based on environmental energy harvesting," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC) Proc.*, May 2015, pp. 266–271.
[69] S. Wang, J. Bi, J. Wu, A. V. Vasilakos, and Q. Fan, "VNE-TD: A virtual network embedding algorithm based on temporal-difference learning," *Comput. Netw.*, vol. 161, pp. 251–263, Oct. 2019.
[70] M. Huang, A. Liu, N. N. Xiong, T. Wang, and A. V. Vasilakos, "An effective service-oriented networking management architecture for 5G-enabled Internet of Things," *Comput. Netw.*, vol. 173, May 2020, Art. no. 107208.
[71] M. T. Lazarescu, "Design of a WSN platform for long-term environmental monitoring for IoT applications," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 3, no. 1, pp. 45–54, Mar. 2013.
[72] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A survey on concept drift adaptation," *ACM Comput. Surv.*, vol. 46, no. 4, pp. 1–37, Apr. 2014.

Martha Arbayani Zaidan (Member, IEEE) received the Ph.D. degree in automatic control and systems engineering from the University of Sheffield, U.K., in 2014. He was a Postdoctoral Research Associate with Maryland University and a Fellow of Aalto University. He is a Postdoctoral Researcher with the Institute for Atmospheric and Earth System Research (INAR), University of Helsinki. His research interests include artificial intelligence and machine learning for intelligent control systems, health monitoring technologies, applied physics, atmospheric, and environmental sciences.

Naser Hossein Motlagh received the D.Sc. degree in networking technology from the School of Electrical Engineering, Aalto University, Finland, in 2018. He is a Postdoctoral Researcher with the Department of Computer Science, Helsinki Institute for Information Technology (HIIT), University of Helsinki. His research interests include the Internet of Things, wireless sensor networks, environmental sensing, and unmanned aerial and underwater vehicles.

Pak L. Fung is currently pursuing the Ph.D. degree with the Institute for Atmospheric and Earth System Research (INAR), University of Helsinki. His Ph.D. work focuses on the derivation of air quality parameters proxy in urban regions. His research interests include remote sensing and machine learning techniques.

David Lu received the B.S. degree in atmospheric science from the University of California at Berkeley, Berkeley. He is a Co-Founder and the CEO of Clarity Movement Company, an environmental tech startup focused on empowering the world to reduce air pollution by assisting municipalities and industries to better measure air quality and take action. Prior to starting Clarity, he was a Key Leader with the University of California's Fossil Fuel Divestment Campaign and worked alongside the Arctic Campaign Director at Greenpeace. Fueled by his everlasting passion for improving our environment, namely mitigating climate change, he believes in using data-driven approaches to address the issues arising in global environmental crises.

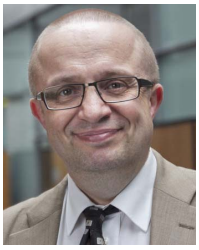
Hilkka Timonen received the Ph.D. degree from the University of Helsinki in 2011 and the Docent title from Tampere University in 2016. She is a Senior Research Scientist and the Head of the Aerosol Composition Group, Finnish Meteorological Institute (FMI). Her research interests include atmospheric aerosol, organic aerosol, black carbon, aerosol mass spectrometry, and emissions measurements

Joel Kuula received the M.Sc. and D.Sc. degrees in mechanical engineering from Aalto University in 2017 and 2020, respectively. He is a Researcher with the Finnish Meteorological Institute (FMI). His research interests include air quality, aerosol, low-cost sensors, mechatronics, and internal combustion engines.

Jarkko V. Niemi received the Ph.D. degree in environmental sciences from the University of Helsinki in 2007. He is an Expert in air quality with the Helsinki Region Environmental Services Authority (HSY). His research interests include urban air quality, physico-chemical characterization and source identification on particles, air quality monitoring methods, and air pollution mitigation measures.

Sasu Tarkoma (Senior Member, IEEE) received the Ph.D. degree in computer science from the University of Helsinki in 2006. He is a Professor of Computer Science with the University of Helsinki and the Head of the Department of Computer Science. He is also affiliated with the Helsinki Institute for Information Technology (HIIT). His research interests include mobile computing, Internet technologies, and AI. He is also a member of the Finnish Union of University Professors and a Professional Member of the Association for Computing Machinery (ACM).

Tuukka Petäjä received the Ph.D. degree and Docent title in physics from the University of Helsinki in 2006 and 2011, respectively. He was a Postdoctoral Researcher with the U.S. National Center for Atmospheric Research (NCAR), Boulder, CO, USA. He is a Professor of Experimental Atmospheric Sciences with the Institute for Atmospheric and Earth System Research (INAR), University of Helsinki. He and his team are currently in charge of the development of aerosol particles measuring equipment for continuous measurements. His research interest includes atmospheric aerosol particles and their role in climate change and air quality.



Markku Kulmala received the Ph.D. degree in theoretical physics from the University of Helsinki in 1988. He is an Academy Professor and the Head of Institute for Atmospheric and Earth System Research (INAR), University of Helsinki. He is the Founder of the International (Station for Measuring Ecosystem-Atmosphere Relations) SMEAR observation networks. He has published more than 800 SCI articles (including more than 40 articles in *Science* and *Nature*). His research interests include atmospheric aerosol nucleation and growth mechanisms, kinetics of atmospheric aerosols and clusters, and biosphere-aerosol-cloud-climate interactions. He is currently a member of the Academy of Europe and a Foreign Academician of the Chinese Academy of Sciences.

Tareq Hussein received the Ph.D. degree in physics from the Division of Atmospheric Sciences in 2005 and the Docent title in physics from the University of Helsinki in 2008. He is a Professor of Physics from the University of Jordan and a Visiting Professor with the Institute for Atmospheric and Earth System Research (INAR), University of Helsinki. His research interests include urban and indoor air quality and exposure.